

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
8 February 2001 (08.02.2001)

PCT

(10) International Publication Number  
WO 01/09384 A2

(51) International Patent Classification<sup>7</sup>: C12Q 1/68 (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.

(21) International Application Number: PCT/US00/20557 (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(22) International Filing Date: 28 July 2000 (28.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/146,437 29 July 1999 (29.07.1999) US

(71) Applicant (*for all designated States except US*): GENZYME CORPORATION [US/US]; Legal Department, Metrowest Place, 4th floor, 15 Pleasant Street Connector, Framingham, MA 01701-9322 (US).

(72) Inventor; and  
(75) Inventor/Applicant (*for US only*): STEFANO, James, E. [US/US]; 199 Saddle Hill Road, Hopkinton, MA 01748 (US).

(74) Agents: KONSKI, Antoinette, F. et al.; Baker & McKenzie, 660 Hansen Way, Palo Alto, CA 94304 (US).



WO 01/09384 A2

Published:  
— *Without international search report and to be republished upon receipt of that report.*

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SERIAL ANALYSIS OF GENETIC ALTERATIONS

(57) Abstract: Methods and compositions for the rapid detection and identification of multiple genetic alterations are provided.

**SERIAL ANALYSIS OF GENETIC ALTERATIONS****TECHNICAL FIELD**

5        The invention is in the field of genetic analysis, including mutation detection and single nucleotide polymorphism (SNP) analysis.

**BACKGROUND**

Nucleotide sequence polymorphism is a hallmark of the human genome. It is  
10    estimated that approximately 0.1% of the 3 billion base-pair human genome is  
subject to polymorphism. Thus, approximately 3 million base pairs of the human  
genome are subject to variation from one individual to another. A unique genetic  
fingerprint can be obtained for each individual, based upon which of these sites  
varies in a given individual, as well as the nature of the variations. The  
15    determination of the constellation of single nucleotide polymorphisms (SNPs) that  
exist in the genome of a given individual will be useful for disease phenotyping, as  
molecular markers for genome mapping, and as markers for forensic identification  
and related techniques. SNPs are but one form of the genetic variation that exists  
within the human population. Other forms of genetic variation include insertions and  
20    deletions of nucleotide sequences, sequence repetitions, translocations and  
inversions.

It is now clear that genetic variation contributes to many, if not most, types of  
human disease. Genetic factors can affect an individual's susceptibility to disease, as  
well as the response of an individual to pharmaceuticals used in the treatment of  
25    disease. In addition, both inherited and acquired genetic changes contribute to the  
development and progression of major human diseases such as cancer, cardiovascular  
disease, neurological and neurodegenerative disease. Consequently, the unique  
nature of an individual's genetic variation will be informative with respect to that  
individual's susceptibility to disease and response to treatment.

30        Most diseases result from the cumulative effect of multiple genetic changes.  
Hence, methods for the parallel analysis of multiple mutations will contribute greatly

to our understanding of the molecular mechanisms of human disease, and to our eventual ability to design more effective treatments. In addition, the effectiveness of a forensic identification increases with the number of markers that are tested. Hence, the ability to conduct parallel analyses of multiple genetic changes will facilitate 5 progress in these and other areas.

Although there are several existing methods for the identification of genetic polymorphisms, none of them allow rapid characterization and parallel analysis of multiple genetic changes. Many methods identify the presence of the mutation, but do not identify the mutation. However, this method does not lend itself easily to 10 analysis of multiple sequence changes, nor is it certain that every sequence difference will result in a corresponding conformational change. Restriction fragment length polymorphism (RFLP) analysis will only detect sequence changes that occur in a restriction enzyme recognition site and, hence, its usefulness is limited. Array methods for analysis of genetic variation involve hybridization to oligonucleotide 15 arrays which contain probes complementary to both the wild-type sequence and the polymorphic variant(s). However, array methods require construction of the arrays and often require elaborate controls for distinguishing single-nucleotide differences in sequence.

U.S. Patent No. 5,459,039 discloses a method for detecting base sequence 20 differences between two DNA molecules, using a protein that recognizes base pair mismatches, by detection of a protein: DNA complex. U.S. Patent No. 5,459,039 does not provide a method for parallel analysis of multiple genetic alterations, nor does it provide a method for identifying genetic alterations.

U.S. Patent No. 5,695,937 discloses a method for serial analysis of gene 25 expression. It does not describe methods for detection and identification of genetic alterations, nor does it disclose methods for parallel analysis of multiple genetic alterations.

A method for rapid parallel analysis of multiple genetic alterations would be particularly useful, in light of the vast genetic diversity of the human species, the 30 consequent preponderance of genetic alterations in the human genome, and the

importance of these genetic variations for diagnosis and treatment of disease, among other things. This invention provides this method.

#### DISCLOSURE OF THE INVENTION

5        The present invention provides methods and compositions for the rapid parallel analysis of multiple genetic alterations in a collection of sample nucleic acids. Practice of the invention allows the identification of one or more genetic alterations in one or more sample polynucleotides, and in addition provides the nucleotide sequences of the genetic alterations so identified.

10       In one embodiment, the invention provides a method for determining the nucleotide sequence of a sample polynucleotide containing one or more genetic alterations.

      In practicing this method, the following steps are performed:

15       (a) contacting a duplex containing the sample polynucleotide and a reference polynucleotide with a reagent which recognizes a non-duplex polynucleotide structure, to form a reagent-heteroduplex complex, wherein the duplex contains at least one base-pair mismatch;

      (b) contacting the reagent-heteroduplex complex with a nucleolytic agent to produce a protected fragment;

20       (c) joining a double-stranded adapter oligonucleotide to the protected fragment;

      (d) amplifying the products of step (c) using primers complementary to a portion of the adapter oligonucleotide, to generate amplification products;

25       (e) joining the amplification products to one another to form a concatemer, wherein each monomeric unit of the concatemer comprises a region of the sample polynucleotide containing the genetic alteration, and the monomeric units are separated by regions of sequence corresponding to a portion of the adapter oligonucleotide sequence; and

      (f) determining the nucleotide sequence of the concatemer.

30       In one embodiment of the invention, multiple sample polynucleotides are assayed using the same reference polynucleotide. In additional embodiments, a

plurality of reference polynucleotides are used to assay one or more sample polynucleotides.

The invention additionally provides oligonucleotide adapters useful for forming concatemers of sequences containing genetic alterations. The 5 oligonucleotide adapters of the invention are designed to minimize self-ligation, to be resistant to nucleolytic activities and to comprise a capture moiety that allows selective immobilization and/or removal of one strand of an amplification product.

Methods and compositions for selective amplification of one strand of a duplex are also provided.

10

#### BRIEF DESCRIPTION OF THE FIGURES

Figures 1A and 1B are a schematic diagram of the assay method. In Figure 1A, reference and test DNAs are heteroduplexed by heating and slow cooling. For each nucleotide difference, two mismatched nucleotide pairs are generated (only one 15 is shown in the figure). These heteroduplexes are contacted with mutS protein. Digestion of the complexed DNA results in protection of small mostly double-stranded DNA fragments bound to mutS. These are purified from the other digest products and DNase by spin column chromatography. The ends of the protected fragments are polished with T4 DNA polymerase.

20 Figure 1B is a continuation of Figure 1A. A pair of double-stranded adapters is ligated to the ends of the protected, polished fragments using T4 DNA ligase. The adapters lack 5' phosphates, leaving one strand of the adapter unligated. Contact with an exonuclease (T7 gene 6 exonuclease) destroys the unligated strand allowing extension of the 3' ends of the protected fragment to be extended by T4 DNA 25 polymerase. The product is amplified by PCR using a set of biotinylated primers corresponding to one strand of the adapters. The PCR products are digested with a restriction enzyme (Hind III shown), the terminal subfragments removed by capture on streptavidin agarose, and the internal subfragment (containing the protected fragment sequence) ligated together in tandem arrays, and then ligated into a vector.

30 Figures 2A to 2C show concerted generation of tandem arrays from amplified protected fragments.

In Figure 2A, a plasmid vector bearing nonpalindromic single-stranded overhangs is generated by ligation of two adapter oligonucleotides complimentary to the restriction fragment ends used to linearize the vector. Both adapters have identical sequence single-stranded overhangs on the non-ligating terminus, except 5 that one of the two lacks a 5' phosphate. The adapter-decorated vector is then purified from any free adapter.

Figure 2B shows the restricted, and purified amplified protected fragment population is ligated together in the presence of a small amount (5%) of a chain-terminating adapter oligonucleotide complimentary on one end to the ends of the 10 fragments. The other end of this adapter is complementary to the single-stranded overhang on the decorated vector. The mixture is ligated to completion. All chains receiving a chain-terminator will be forced to form larger linear arrays until another chain terminator (or another growing chain also ending in a terminator) is ligated yielding a large tandem array whose average size is determined by the ratio of 15 terminator to fragment ratio.

Figure 2C shows the decorated vector is added to the ligation reaction to accept the array. To minimize repair activity directed against the final ligation product, one strand may be removed by T7 gene 6 exonuclease.

20

#### MODES FOR CARRYING OUT THE INVENTION

The practice of the present invention employs, unless otherwise indicated, conventional techniques of microbiology, molecular biology, recombinant DNA and related fields, which are within the skill of the art. These techniques are fully explained in the literature. See, *e.g.*, Maniatis et al., *Molecular Cloning: A 25 Laboratory Manual* (1982) Cold Spring Harbor Laboratory Press; D. Glover, ed., *DNA Cloning: A Practical Approach*, vols. I & II (1985) IRL Press; M. Gait, ed., *Oligonucleotide Synthesis* (1984) IRL Press; B. Harnes & S. Higgins, eds., *Nucleic 30 Acid Hybridization* (1985) IRL Press; Perbal, *A Practical Guide to Molecular Cloning* (1984); Ausubel, et al., *Current Protocols In Molecular Biology* (1987 and annual updates) John Wiley & Sons; and Sambrook et al., *Molecular Cloning: A*

Laboratory Manual (2nd Edition); vols. I, II & III (1989) Cold Spring Harbor Laboratory Press.

### **Definitions**

5        As used herein, certain terms will have specific meanings. The singular form "a," "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

10      The term "comprising" is intended to mean that the compositions and methods include the recited elements, but not excluding others. "Consisting essentially of" when used to define compositions and methods, shall mean excluding other elements of any essential significance to the combination. Thus, a composition consisting essentially of the elements as defined herein would not exclude trace contaminants from the isolation and purification method and pharmaceutically acceptable carriers, such as phosphate buffered saline, preservatives, and the like. "Consisting of" shall mean excluding more than trace elements of other ingredients and substantial method steps for administering the compositions of this invention. Embodiments defined by each of these transition terms are within the scope of this invention.

15      The terms "polynucleotide" and "nucleic acid molecule" are used interchangeably to refer to polymeric forms of nucleotides of any length. The polynucleotides may contain deoxyribonucleotides, ribonucleotides, and/or their analogs. Nucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The term "polynucleotide" includes single-, double-stranded and triple helical molecules. "Oligonucleotide" refers to polynucleotides of between about 6 and about 100 nucleotides of single- or double-stranded DNA or RNA. Oligonucleotides are also known as oligomers and may be isolated from genes, or chemically synthesized by methods known in the art. A "primer" refers to an oligonucleotide, usually single-stranded, that provides a 3'-hydroxyl end for the initiation of nucleic acid synthesis.

The following are non-limiting embodiments of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A nucleic acid molecule may also comprise modified nucleic acid molecules, such as methylated nucleic acid molecules and nucleic acid molecule analogs. Analogs of purines and pyrimidines are known in the art, and include, but are not limited to, aziridinycytosine, 4-acetylcytosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, pseudouracil, 5-pentylnyluracil and 2,6-diaminopurine. The use of uracil as a substitute for thymine in a deoxyribonucleic acid is also considered an analogous form of pyrimidine.

Oligonucleotides are short polymers of nucleotides, generally less than 200 nucleotides, preferably less than 150 nucleotides, more preferably less than 100 nucleotides, more preferably less than 50 nucleotides and most preferably less than 30 nucleotides in length. Oligonucleotides are generally considered to comprise shorter polymers of nucleotides than do polynucleotides, although there is an art-recognized overlap between the upper limit of oligonucleotide length and the lower limit of polynucleotide length. Consequently, for the purposes of the present invention, the terms "oligonucleotide" and "polynucleotide" shall not be considered limiting with respect to polymer length.

As used herein, "base pair," also designated "bp" refers to the complementary nucleic acid molecules; in DNA the purine adenine (A) is hydrogen bonded with the pyrimidine base thymine (T), and the purine guanine (G) with pyrimidine cytosine (C), also known as Watson-Crick base-pairing. A thousand base pairs is often called a kilobase, or kb. A "base pair mismatch" refers to a location in a nucleic acid molecule in which the bases are not complementary Watson-Crick pairs.

The term "duplex" refers to the complex formed between two strands of hydrogen-bonded, complementary nucleic acid molecules. A duplex need not be

entirely complementary, but can contain one or more mismatches or one or more deletions or additions. A duplex is sufficiently long-lasting to persist between formation of the duplex or complex and subsequent manipulations, including, for example, any optional washing steps.

5 As used herein, the term "reference strand" or "wild-type strand" refers to the nucleic acid molecule or polynucleotide having a sequence prevalent in the general population that is not associated with any disease or discernible phenotype. It is noted that in the general population, wild-type genes may include multiple prevalent versions that contain alterations in sequence relative to each other and yet do not cause a discernible pathological effect. These variations are designated "polymorphisms" or "allelic variations." It is therefore possible to prepare multiple reference strands, thereby providing a mixture of the most common polymorphisms. Alternatively, one reference strand may be used that has been selected for its particular sequence. The reference strand can also be chemically or enzymatically 10 modified, for example to remove or add methyl groups. In one or more embodiments, the reference strand is comprised of a PCR product identical at least in part to the sequence prevalent in the general population.

15

In a preferred embodiment, the reference strand or wild-type strand comprises a portion of a particular gene or genetic locus in the patient's genomic DNA known 20 to be involved in a pathological condition or syndrome. Non-limiting examples of genetic syndromes include cystic fibrosis, sickle-cell anemia, thalassemias, Gaucher's disease, adenosine deaminase deficiency, alpha1-antitrypsin deficiency, Duchenne muscular dystrophy, familial hypercholesterolemia, fragile X syndrome, glucose-6-phosphate dehydrogenase deficiency, hemophilia A, Huntington's disease, 25 myotonic dystrophy, neurofibromatosis type 1, osteogenesis imperfecta, phenylketonuria, retinoblastoma, Tay-Sachs disease, and Wilms tumor (Thompson and Thompson, *Genetics in Medicine*, 5th Ed.).

In another embodiment, the reference strand comprises part of a particular gene or genetic locus that may not be known to be linked to a particular disease, but 30 in which polymorphism is known or suspected. For example, obesity may be linked with variations in the apolipoprotein B gene, hypertension may be due to genetic

variations in sodium or other transport systems, aortic aneurysms may be linked to variations in I-haptoglobin and cholesterol ester transfer protein, and alcoholism may be related to variant forms of alcohol dehydrogenase and mitochondrial aldehyde dehydrogenase. Furthermore, an individual's response to medicaments may be 5 affected by variations in drug modification systems such as cytochrome P450s, and susceptibility to particular infectious diseases may also be influenced by genetic status. Finally, the methods of the present invention can be applied to HLA analysis for identity testing.

The term "sample strand" or "patient strand" refers to the polynucleotide 10 having unknown sequence and potentially containing one or more mutations or mismatches as compared to the reference strand. This may be a PCR product amplified from patient DNA or other sample(s).

In yet another embodiment, the reference strand comprises part of a foreign 15 genetic sequence *e.g.*, the genome of an invading microorganism. Non-limiting examples include bacteria and their phages, viruses, fungi, protozoa, myoplasms, and the like. The present methods are particularly applicable when it is desired to distinguish between different variants or strains of a microorganism in order to choose appropriate therapeutic interventions.

The term "genetic alterations" or "mutations" is used to refer to a change 20 from the wild-type or reference sequence of one or more nucleic acid molecules. It refers to base pair substitutions, additions and deletions of a sample strand when compared to a reference strand.

A linear sequence of polynucleotides is "substantially homologous" to 25 another linear sequence, having the opposite polarity, if both sequences are capable of hybridizing to form duplexes with the same complementary polynucleotide. Sequences that hybridize under conditions of high stringency are more preferred. It is understood that hybridization reactions can accommodate insertions, deletions, and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely 30 correspond or align. Preferably, the "substantially homologous" sample sequences of

the invention contain a single mutation (mismatch) or an addition or deletion of 1 to about 10 base pairs when compared to the reference polynucleotide.

As used herein, the term “reagent which recognizes a non-duplex polynucleotide structure” is any agent, proteinaceous or otherwise, which provides 5 this functional activity when used in the method of this invention. In one embodiment, this agent is a mismatch binding protein or “MBP.” MBP refers to the group of proteins which recognize and bind to nucleotide mismatches and unpaired nucleotides in polynucleotide duplexes. As used herein, the term “non-duplex polynucleotide structure” shall mean the absence at any position of a Watson-Crick 10 base pair, *i.e.*, any pair other than A:T or G:C, or A:U in RNA, and unpaired nucleotides. By recognizing and binding to improperly paired nucleotide strands, these proteins are involved in the complex pathway of genetic repair. Repair is generally initiated by the binding of the protein mutS to the mismatch. (*See*, Modrich (1994), *supra*). According to current understanding, the portion of DNA 15 between the mutS-bound mismatch and the nearest GATC element (bound by mutH) is looped out by a translocase activity of the mutS protein, assisted by the DNA helicase activity of mutL, leading to activation of the GATC endonuclease associated with mutH. Cooperative action of mutS, mutH and the DNA helicase (mutL) is required to mark the mismatch region, which is then repaired using exonucleases and 20 polymerases.

“MBPs” includes several embodiments. These embodiments include any fragment, analog, mutein, variant or mixture thereof, that retains the ability to recognize and bind to a nucleotide mismatch. In one embodiment, a “variant” is a protein or polypeptide with conservative amino acid substitutions as compared to the 25 wild-type amino acid sequence: The term therefore encompasses MutS and its homologues including hMSH2, hPMS1, and hPMS2.

Mismatch repair proteins for use in the present invention may be derived from *E. coli* (as described above) or from any organism containing mismatch repair proteins with appropriate functional properties. Non-limiting examples of useful 30 proteins include those derived from *Salmonella typhimurium* (MutS, *see*, Su, S.S. and Modrich, P., *Proc. Natl. Acad. Sci. USA* 84:5057-5061 (1986); MutL); *Streptococcus*

*pneumoniae* (HexA, HexB); *Saccharomyces cerevisiae* ("all-type," MSH2, MLH1, MSH3); *Schizosaccharomyces pombe* (SWI4); mouse (rep1, rep3); and human ("all-type," hMSH2, hMLH1, hPMS1, hPMS2, duc1). In another embodiment, heteroduplexes formed between patients' DNA and wild-type DNA as described above are incubated with p53 or its C-terminal domain (Lee, et al., *Cell* 81:1013-1020 (1995)). In another embodiment, purified MutS, MutL, and MutH are used to cleave mismatch regions (Su et al., *Proc. Natl. Acad. Sci. USA* 83:5057 (1986); Grulley et al., *J. Biol. Chem.* 264:1000 (1989)).

When the agents are proteins or polypeptides, they can be in the L or D form so long as the biological activity of the polypeptide is maintained. For example, the protein can be altered so as to be secreted from the cell for recombinant production and purification. These also include proteins which are post-translationally modified by reactions that include glycosylation, acetylation and phosphorylation. Such polypeptides also include analogs, alleles and allelic variants which can contain amino acid derivatives or non-amino acid moieties that do not affect the biological or functional activity of the protein as compared to wild-type or naturally occurring protein. The term amino acid refers both to the naturally occurring amino acids and their derivatives, such as TyrMe and PheCl, as well as other moieties characterized by the presence of both an available carboxyl group and an amine group. Non-amino acid moieties which can be contained in such polypeptides include, for example, amino acid mimicking structures. Mimicking structures are those structures which exhibit substantially the same spatial arrangement of functional groups as amino acids but do not necessarily have both the  $\alpha$ -amino and  $\alpha$ -carboxyl groups characteristic of amino acids.

As used herein, the term "nucleolytic agent" refers to an enzyme or chemical agent that cleaves at least one strand of DNA. Non-limiting examples of such agents include exonucleases such as BAL31, lambda exonuclease, exonuclease III, T7 gene 6 exonuclease, and endonucleases such as DNase I, *S. aureus* (micrococcal) nuclease, P1 nuclease, and the like. In addition, examples of chemical agents include bleomycin and/or iron-bound intercalators such as 0-phenanthroline that can direct

the formation of hydroxyl radicals in close proximity to a DNA-bound intercalator to effect DNA cleavage.

The term "exonuclease" refers to an enzyme that cleaves nucleotides sequentially from the free ends of a linear nucleic acid substrate. Exonucleases can 5 be specific for double or single stranded nucleotides and/or directionally specific, for instance, 3'→5' and/or 5'→3'. Some exonucleases exhibit other enzymatic activities, for example, native T7 DNA polymerase is both a polymerase and an active 3'→5' exonuclease. Exonuclease III removes nucleotides one at a time from the 3'-end of duplex DNA, exonuclease VII makes oligonucleotides by several 10 nucleotides from both ends of single-stranded DNA and lambda exonuclease removes nucleotides having attached 5' phosphate groups from the 5' end of duplex DNA.

As used herein, the term "nick translation" refers to a process comprising the combined action of a 5'→3' exonuclease and a polymerase to degrade the 5' 15 terminus of a polynucleotide in duplex with a template strand polynucleotide and extend the 3' terminus of an adjacent second polynucleotide strand in duplex with the same template strand polynucleotide wherein the said 5' terminus is separated from the 3' terminus of the adjacent polynucleotide by a nick or gap. For these purposes, "adjacent" includes any number of nucleotides without limit (more than zero 20 nucleotides is usually referred to as a "gap"), but is typically zero nucleotides (usually referred to as a "nick"). As is well understood in the art, such exonuclease and polymerase activities may reside in the same enzyme (such as with *E. coli* DNA polymerase I) or different enzymes, and their action may be either simultaneous or sequential.

25 The term "polymerase chain reaction" or "PCR" refers to a method for amplifying a DNA base sequence using a heat-stable polymerase such as Taq polymerase, and two oligonucleotide primers, one complementary to the (+)-strand at one end of the sequence to be amplified and the other complementary to the (-)-strand at the other end. Because the newly synthesized DNA strands can 30 subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation can produce rapid

and highly specific exponential amplification of the desired sequence. PCR also can be used to detect the existence of the defined sequence in a DNA sample.

This invention provides a method for determining the sequence of a sample polynucleotide containing one or more genetic alterations. The genetic alteration are 5 presented in the form of a heteroduplex of polynucleotides which is then contacted with a reagent that recognizes a non-duplex polynucleotide structure, to form a reagent-heteroduplex complex, wherein the duplex contains at least one base-pair mismatch. As used herein, a heteroduplex can comprise: DNA:DNA; DNA:RNA; or DNA:RNA. A nucleolytic agent is contacted with the heteroduplex to produce a 10 protected fragment from which a double-stranded adapter oligonucleotide is joined. The duplex:adapter complex is amplified using primers complementary to a portion of the adapter oligonucleotide, to generate amplification products. The products are the amplification products joined to one another to form a concatemer, wherein each monomeric unit of the concatemer comprises a region of the sample polynucleotide 15 containing the genetic alteration or the substantially-complementary reference strand region, and the monomeric units are separated by regions of sequence corresponding to a portion of the adapter oligonucleotide sequence. The sequence of the concatemer is comprised of sequences of the strands of the protected fragments in random order and orientation. The sequence or sequences of the protected fragment 20 strands are compared with the known sequence of the sample strand to identify the genetic alterations. Alternatively, in cases where the sequence of the reference strand is not fully known, the sequence of the two strands of the same protected fragment and/or sequences of other protected fragments arising from the same mismatch (identifiable as comprising substantially-identical overlapping or nested sequences) 25 are compared to each other to identify the nucleotide sequence variations within or between the sample and/or reference polynucleotides. In one embodiment, a plurality of reference polynucleotides are used with either the same or different sample strands.

In one aspect, the adapter oligonucleotide comprises one or more restriction 30 enzyme recognition sites. In a further aspect, the amplification products are digested with a restriction enzyme which cleaves at the recognition site prior to joining to

adapter polynucleotides. In a yet further aspect, adapter polynucleotides of differing sequences are utilized to join the amplified products to each other.

In a further embodiment, the adapter oligonucleotide has the following characteristics: It contains an inner end and an outer end; it is non-phosphorylated at 5 the 5' terminus of the inner end; it contains a capture moiety at the outer end, and it has one or more blocking linkages adjacent to the capture moiety. In a further embodiment, joining of the adapter comprises joining the first strand of the double-stranded adapter oligonucleotide to the products of the prior step and joining the second strand of the double-stranded adapter oligonucleotide to the products of this 10 step. As set forth in more detail below, a further embodiment of the invention provides the first strand is covalently joined to the protected fragment by ligation, and the second strand is covalently joined to the protected fragment by nick-translation.

It is to be understood, although not always explicitly stated, that the methods 15 as disclosed herein can be practiced with a plurality of sample strands with one or more reference strands, and vice versa. In addition, a plurality of reference strands can be initially contacted with a plurality of sample strands.

#### Materials and Methods

##### 20 **Preparation of Sample and Reference Polynucleotides**

Reference DNA can be synthesized by chemical means or, preferably, isolated from any organism by any method known in the art. The organism will have no discernible disease or phenotypic effects. This DNA may be obtained from any cell source, tissue source or body fluid. Non-limiting examples of cells sources 25 available in clinical practice include blood cells, buccal cells, cerviovaginal cells, epithelial cells from urine, fetal cells, or any cells present in tissue obtained by biopsy. Body fluids include urine, blood, cerebrospinal fluid (CSF), and tissue exudates at the site of infection or inflammation. DNA is extracted from the cells or body fluid using any method known in the art. Preferably, at least 5 pg of DNA is 30 extracted. The extracted DNA can be used without further modification or stored for future use.

Preferably, one or more specific regions in the extracted reference polynucleotide are amplified by PCR using a set of PCR primers complementary to genomic DNA separated by up to about 500 base pairs. PCR conditions found to be suitable are described below in the Examples. It will be understood that optimal

5 PCR conditions can be readily determined by those skilled in the art. (See, e.g., PCR 2: A PRACTICAL APPROACH (1995) eds. M.J. McPherson, B.D. Hames and G.R. Taylor, IRL Press, Oxford).

PCR products can be purified by a variety of methods, including but not limited to, microfiltration, dialysis, gel electrophoresis and the like. It is desirable to 10 remove the polymerase used in PCR so that no new DNA synthesis can occur.

### Duplex Formation

A reference:sample heteroduplex can be formed by any method of hybridization known in the art. In one embodiment, the reference and samples are 15 separately heated and then annealed together. Preferably the heating step is between about 70°C and about 100°C, more preferably between about 80°C and 100°C, and even more preferably between about 90°C and 100°C. The polynucleotide is kept at the elevated temperature for sufficient time to separate the strands, preferably between about 2 minutes and about 15 minutes, more preferably between about 2 and 20 about 10 minutes and even more preferably about 5 minutes.

The separately heated reference and sample strands are then combined while at the elevated temperatures and allowed to cool. Generally, cooling occurs rather slowly, for instance the solution is allowed to cool to 50°C over a period of about an hour. The cooling must be sufficiently slow as to allow formation of 25 reference:sample duplexes including those with both high and low Tm. The duplexes can be used immediately, or stored at 4°C until use.

Alternatively, a duplex can be formed by adjusting the salt and temperature to achieve suitable hybridization conditions. Hybridization reactions can be performed in solutions ranging from about 10 mM NaCl to about 600 mM NaCl, at 30 temperatures ranging from about 37°C to about 65°C. It will be understood that the stringency of the hybridization reaction is determined by both the salt concentration

and the temperature. For instance, a hybridization performed in 10 mM salt at 37°C may be of similar stringency to one performed in 500 mM salt at 65°C. In addition, organic solvents and/or chaotropic salts such as guanidine thiocyanate (2.5M) may be used, allowing hybridization to be performed at 37°C. Finally, means of accelerating 5 hybridization such as phenol emulsion (Miller & Riblet, *Nucl. Acids Res.* 23:2339-2340 (1995)) can be employed. For the present invention, any hybridization conditions can be used that form hybrids between substantially homologous complementary sequences, provided the reagents employed are compatible with the MBP and exonuclease employed. Generally, this can be accomplished by exchange 10 into the reaction buffer of choice by dilution, extraction followed by ethanol precipitation, ultrafiltration or spin column chromatography and the like. In a preferred embodiment stringent hybridization conditions are used.

A genetic alteration is a difference in nucleotide sequence between two polynucleotides. In certain cases, a genetic alteration can be a mutation, resulting in 15 a detectable phenotype. In other cases, a genetic alteration may not be linked to a phenotype, but will be useful, *e.g.*, for forensic purposes. A particular sample polynucleotide that is assayed in the practice of the invention can contain a single genetic alteration or it may contain multiple genetic alterations. In a preferred embodiment, the genetic alteration will be a single nucleotide polymorphism, *i.e.*, a 20 change in the sequence of a single base, compared to the wild-type sequence. In a preferred embodiment, a plurality of sample polynucleotides, each of which contains one or more genetic alterations, is analyzed by the practice of the invention, and the nucleotide sequences of the genetic alterations thereby determined.

To identify and determine the sequences of genetic alterations, one or more 25 sample polynucleotides are hybridized with one or more reference polynucleotides to generate a polynucleotide duplex. A polynucleotide duplex is a double-stranded polynucleotide, in which the association between the two strands is mediated, at least in part, by complementary base-pairing. Hybridization of polynucleotides to form duplexes proceeds according to well-known and art-recognized base-pairing 30 properties, such that adenine base-pairs with thymine or uracil, and guanine base-pairs with cytosine. The property of a nucleotide that allows it to base-pair with a

second nucleotide is called complementarity. Thus, adenine is complementary to both thymine and uracil, and *vice versa*; similarly, guanine is complementary to cytosine and *vice versa*. In a polynucleotide duplex, one or both of the two component strands may not be duplex along their entire length if, for example, one 5 strand is longer than the other, or if the two strands have non-complementary terminal sequences.

For the purposes of the present invention, it is useful to distinguish between a homoduplex, in which all bases in both component polynucleotide strands form complementary base pairs (along the double-stranded portion of the duplex) and a 10 heteroduplex, in which one or more additions, deletions or base-pair mismatches exist within the duplex. Formation of a heteroduplex is indicative of a genetic alteration in one of the two polynucleotide strands of a duplex, with respect to the other strand. Two polynucleotide strands containing one or more additions, deletions or mismatches with respect to each other are nevertheless capable of forming a stable 15 duplex, if sufficient complementary exists between the two strands. Moreover, those of skill in the art are aware that lower hybridization stringency allows formation of stable duplexes between polynucleotide strands with higher degrees of noncomplementarity. Therefore, hybridization conditions can be adjusted to facilitate the formation of heteroduplexes. For the purposes of the invention, the 20 stability of a heteroduplex is such that it will persist after its formation through subsequent manipulations such as washing, protein binding and treatment with nucleolytic agents.

The nucleotide sequence of a reference polynucleotide will be known as a reference sequence and, in general, a reference sequence will be the wild-type 25 sequence of a particular genetic region. Variations from the wild-type sequence present in the sample polynucleotide(s) can thereby be determined. In one embodiment of the invention, a single reference sequence, comprising a region of known genetic variability, is used to assay multiple sample polynucleotides. In additional embodiments, a reference polynucleotide could comprise a sequence that 30 has been defined as "mutant" and be used to detect either wild-type sequences or sequences containing a different mutant allele.

In the practice of the invention, a sample polynucleotide, either single- or double-stranded, is contacted with a reference polynucleotide, either single- or double-stranded, to form a mixture. The mixture is treated so as to denature double-stranded polynucleotides and to remove regions of secondary structure in single-stranded polynucleotides, for instance, by heating. After denaturation, the mixture is incubated in solution under conditions of temperature, ionic strength, pH, etc., that are favorable to annealing, *i.e.*, under hybridization conditions. Hybridization conditions are chosen to allow duplex formation between sequences having one or more mismatches, as is known in the art.

10        Hybridization of a sample polynucleotide with a reference polynucleotide, according to the practice of the invention, will thus result in the formation of a heteroduplex polynucleotide, containing one or more mismatches, *i.e.*, sites at which complementary base-pairing does not occur.

It is noted that in the general population, wild-type genes may include multiple prevalent versions that contain alterations in sequence relative to each other and yet do not cause a discernible pathological effect. These variations are designated "polymorphisms" or allelic variations." It is therefore possible to prepare multiple reference polynucleotides, thereby providing a mixture of the most common polymorphisms. Alternatively, a single reference polynucleotide may be used that has been selected for its particular sequence. The reference polynucleotide can also be chemically or enzymatically modified, for example to remove or add methyl groups. In one or more embodiments, the reference polynucleotide comprises a PCR product identical, at least in part, to the sequence prevalent in the general population. It is intended to include, but not be limited to, polynucleotides as defined above, *i.e.*, a gene or gene fragment, restriction fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers.

30        In a preferred embodiment, the reference polynucleotide comprises a portion of a particular gene or genetic locus in the patient's genomic DNA known to be involved in a pathological condition or syndrome. Non-limiting examples of genetic

syndromes include cystic fibrosis, sickle-cell anemia, thalassemias, Gaucher's disease, adenosine deaminase deficiency, alphas-antitrypsin deficiency, Duchenne muscular dystrophy, familial hypercholesterolemia, fragile X syndrome, glucose 6-phosphate dehydrogenase deficiency, hemophilia A, Huntington's disease, myotonic dystrophy, neurofibromatosis type 1, osteogenesis imperfecta, phenylketonuria, retinoblastoma, Tay-Sachs disease, and Wilms' tumor. Thompson and Thompson, *Genetics in Medicine*, 5<sup>th</sup> Ed.

In another embodiment, the reference polynucleotide comprises part of a particular gene or genetic locus that may not be known to be linked to a particular disease, but in which polymorphism is known or suspected. For example, obesity may be linked with variations in the apolipoprotein B gene, hypertension may be due to genetic variations in transport systems for sodium or other ions, aortic aneurysms may be linked to variations in I-haptoglobin and cholesterol ester transfer protein, and alcoholism may be related to variant forms of alcohol dehydrogenase and mitochondrial aldehyde dehydrogenase. Furthermore, an individual's response to medicaments may be affected by variations in drug modification systems such as cytochrome P-450s, and susceptibility to particular infectious diseases may also be influenced by genetic status. Finally, the methods of the present invention can be applied to HLA analysis for identity testing.

20

### Protected Fragments

To facilitate the rapid analysis of multiple genetic alterations, the invention allows the generation of a smaller subregion of the sample polynucleotide:reference polynucleotide heteroduplex, wherein the subregion encompasses the sequence difference between the sample polynucleotide and the reference polynucleotide. In one embodiment, the subregion is generated by protection from nuclease digestion, and is therefore denoted a protected fragment. An exemplary method for generating a protected fragment is as follows.

A heteroduplex formed between a sample polynucleotide and a reference polynucleotide is contacted with a reagent which recognizes a non-duplex structure. The reagent can be chemical or enzymatic. In a preferred embodiment, the reagent is

enzymatic; in a more preferred embodiment, the reagent is a mismatch binding protein; in a still more preferred embodiment, the reagent is the *E. coli* mutS protein. Reaction conditions compatible with the binding of mutS and other mismatch binding proteins to heteroduplexes are known in the art and are additionally provided

5 in the examples below. Those of skill in the art are aware that such conditions can be varied and are also aware of art-recognized methods for the detection of protein binding to a heteroduplex, such as filter binding, gel electrophoresis and the like. Thus, additional conditions compatible with binding can be determined without undue experimentation.

10 Contact between the reagent and the heteroduplex results in the formation of a reagent-heteroduplex complex. This complex will generally comprise a duplex polynucleotide in which one (or more) portions of the duplex are bound by the reagent (in the vicinity of the genetic alteration), and the remainder of the duplex is free of bound reagent. Those portions of the duplex that are free of bound reagent are

15 susceptible to nucleolytic agents. Accordingly, to generate a protected fragment, a reagent:heteroduplex complex is subjected to the action of one or more nucleolytic agents. Such nucleolytic agents can be chemical or enzymatic.

An enzymatic nucleolytic agent is also known as a nuclease. Thus, a nuclease is an enzyme capable of degrading nucleic acids. An exonuclease degrades from the

20 ends of a nucleic acid molecule. A 5'-specific exonuclease will begin degradation at the 5' end of a nucleic acid molecule, and a 3'-specific exonuclease will begin degradation at the 3' end of a nucleic acid molecule. 5'-specific exonucleases may additionally be specific for either 5'-phosphate- or 5'-hydroxyl terminated ends.

Similarly, 3'-specific exonucleases may be specific for either 3'-phosphate- or 3'-

25 hydroxyl terminated ends. An endonuclease degrades internally in a nucleic acid molecule. A single strand-specific nuclease degrades single-stranded nucleic acids, either exonucleolytically or endonucleolytically, but is unable to degrade a double-stranded nucleic acid.

A preferred nuclease is an endonuclease. Suitable endonucleases include S1

30 nuclease, p1 nuclease from Micrococcal nuclease, Mung Bean nuclease and DNase I. Concentrations of endonuclease sufficient to digest non-protected portions of a

duplex polynucleotide are known in the art and an example is provided in the examples, *infra*.

Nucleolytic agents of a chemical nature include, for example, NaOH or other bases, which are capable of nucleolytic degradation of RNA.

5        In another embodiment of the invention, the requirement for a reagent that recognizes a non-duplex polynucleotide structure is fulfilled by the concerted, sequential action of several reagents. For example, a reagent-heteroduplex complex can be contacted with a chemical or enzymatic reagent which cleaves one strand of an otherwise duplex polynucleotide at or near a mismatch, followed by contact of the  
10      nicked polynucleotide with a nick-binding protein. Suitable reagents which cleave at or near a mismatch include S1 nuclease, Mung Bean nuclease, Mut Y protein, and Mut M protein.

15      The region of the polynucleotide component of a reagent-heteroduplex complex that is bound by the reagent is protected from the action of the nucleolytic agent. Thus, treatment of the reagent-heteroduplex complex with a nucleolytic agent generates a protected fragment. Protected fragments can be purified, if desired, by any method which separates the protected fragment from (usually) smaller unprotected polynucleotide fragments produced by the nucleolytic agent. For example, a size separation, such as gel filtration or gel electrophoresis, can be used.

20

### **Mismatch Recognition**

25      The reference:sample duplex is contacted with one or more agents having the ability to specifically bind to bp mismatches. This includes, but is not limited to, mismatch binding proteins. The agent is contacted under conditions which allow binding of the agent to the mismatch. Preferably, the MBP is *E. coli* MutS (AP Biotech) although other MBPs or mixtures of MBPs can be used. For instance, homologues of MutS such as MutS from *Thermus aquaticus* (Epicentre), *Streptococcus pneumoniae* HexA, hMSH2, genetically modified MutS or other mutation binding proteins such as RuvC protein from *E. coli*, human p53, or  
30      genetically modified (non-cleaving forms) of mutY or T4 endonuclease VII may be used. Preferably, the duplex is contacted with MutS at 0°C for between about 10 and

30 min., preferably about 30 min. MutS binding, yielding consistent patterns of protection with a large "footprint" range of bp protected, occurs near neutral pH, preferably between a pH of about 6.5 and 8.5, and more preferably, between a pH of about 7.0 and 8.0. A source of magnesium ions (Mg<sup>++</sup>) can also be added to the 5 reaction to enhance MutS binding.

#### **Adapter oligonucleotides**

To facilitate the formation of tandem arrays of protected fragments (*i.e.*, concatemers) and their sequence analysis, oligonucleotide adapters are added to the protected fragments. The adapters facilitate the amplification of sequences 10 corresponding to a protected fragment, by providing a pair of primer sites. In a preferred embodiment of the invention, an adapter oligonucleotide duplex lacks terminal phosphate residues. The lack of 5'-phosphate termini on the adapter oligonucleotides prevents self-ligation of the adapter oligonucleotides, which would lead to the production of spurious amplification products (*i.e.*, "primer dimers") in 15 later steps. Lack of 5'-phosphate termini also prevents ligation of multiple adapters to a protected fragment, insuring that a single adapter is ligated to each end of a protected fragment.

For the purposes of the invention, it is convenient to distinguish between an inner end and an outer end of the adapter oligonucleotide. The inner end of the 20 adapter oligonucleotide is that end which becomes joined to a protected fragment in the practice of the invention. Conversely, the outer end of the adapter oligonucleotide is that end which is not joined to the protected fragment. Hence, the outer end of the adapter oligonucleotide forms a terminus of the polynucleotide which results from ligation of an adapter oligonucleotide to a protected fragment. It 25 is possible to specify the inner and outer ends of an adapter oligonucleotide in several ways. In one embodiment, the adapter oligonucleotide duplex comprises a capture moiety at its outer end, which serves to allow immobilization of the outer portions of the adapter-modified fragment after digestion with a restriction enzyme, and also serves to sterically block ligation of that end of the adapter oligonucleotide to a 30 protected fragment.

The adapter oligonucleotides can optionally comprise a capture moiety at the outer end of the adapter oligonucleotide duplex. The capture moiety can be attached either to the strand that forms the 5'-end of the outer end of the adapter oligonucleotide duplex, or to the strand that forms the 3'-end of the outer end of the adapter oligonucleotide duplex. The capture moiety is generally a molecule that is capable of interacting with a second molecule (a recognition moiety) to form a stable complex. Adapter oligonucleotides and any other nucleic acid that is directly or indirectly attached to the capture moiety will be present in a complex formed between a capture moiety and a recognition moiety. A recognition moiety will often be attached to a solid substrate, or otherwise immobilized such that a capture moiety:recognition moiety complex can be brought out of solution. Exemplary capture moiety:recognition moiety pairs include biotin:adidin, biotin:streptavidin, biotin:anti-biotin, antigen:antibody, hapten:antibody, enzyme:substrate, sugar:lectin, protein:ligand and nucleic acid:complementary nucleic acid. Other interacting molecules that can serve as capture moiety:recognition moiety pairs will be known to those of skill in the art. It is also clear that the roles of capture moiety and recognition moiety can be reversed.

Adapter oligonucleotides will comprise, within their sequence, a primer binding site. A primer binding site refers to a region of an oligonucleotide or polynucleotide, such as an adapter or a sequence encoded by an adapter, that is capable of base-pairing with a primer, or that encodes a sequence that is able to base-pair with a primer. A primer is an oligonucleotide or polynucleotide capable of base-pairing with another oligonucleotide or polynucleotide and serving as a site from which polymerization can be initiated, normally from a 3'-hydroxyl end. Because of the ease with which oligonucleotides of defined sequence can be synthesized, virtually any sequence, capable of base-pairing, can function as a primer binding site.

An adapter oligonucleotide duplex can comprise one or more blocking linkages adjacent to its outer end. A blocking linkage is an internucleotide linkage which is less susceptible to nucleolytic degradation, compared to the phosphodiester linkage normally found in most naturally-occurring nucleic acids. Exemplary blocking linkages include phosphorothioate, methyl phosphonate, boronate and

others. Other modifications such as nucleic acid analogs (PNAs, morpholidate DNAs, locked nucleoside analogs (LNA), and the like, may also be used. Further, addition of bulky haptens such as fluorescent tags, inverted nucleosides (in 5'-5' or 3'-3' linkage), or biotin, any of which inhibit the action of nucleases used in the 5 procedure, may also be used. The presence of blocking linkages minimizes loss of protected fragments due to nucleolytic action, as described, *infra*.

### **End repair**

In one embodiment of the invention, the protected fragments are rendered 10 blunt-ended (if not already blunt-ended as a result of the action of a nucleolytic agent), prior to the addition of adapters, by one of a number of end repair processes. In one embodiment, this is accomplished by incubating the protected fragments with T4 DNA Polymerase, dATP, dCTP, dGTP and dTTP under conditions suitable for both the polymerization and exonucleolytic activities of T4 DNA polymerase. 15 Fragments containing 3' overhanging ends will be rendered blunt-ended by the 3'-specific exonuclease activity of T4 DNA Polymerase; while 5' overhangs will be converted to blunt ends by polymerization of the recessed 3' termini. In an alternative embodiment, protected fragments are treated with a single strand-specific exonuclease, such as *E. coli* exonuclease VII or exonuclease I. In a preferred 20 embodiment, 5'-phosphate-terminated blunt ends are generated, either through the end repair process or as a result of nucleolytic action. Other methods for generating blunt-ended fragments, such as, for example, physical methods, treatment with single-strand-specific exo- or endonucleases or the use of other types of nucleic acid polymerase, are also within the scope of this invention.

25

### **Joining of adapter oligonucleotides to protected fragments**

Subsequent to end repair of the protected fragments, (a) double-stranded adapter oligonucleotide(s) is/are attached to the protected fragments. In a preferred embodiment, initial attachment of one strand of the adapter oligonucleotide 30 duplex to one strand of a protected fragment is achieved by enzymatic ligation using

a ligase enzyme. Exemplary ligase enzymes include T4 DNA ligase, *E. coli* DNA ligase, *T. aquaticus* ligase.

As a result of either the action of the nucleolytic agent or of the end repair process, the protected fragments will have 5'-phosphate termini. The oligonucleotide adapters, on the other hand, lack terminal phosphate residues. The lack of 5'-phosphate termini on the adapter oligonucleotides prevents self-ligation of the adapter oligonucleotides, which would lead to the production of spurious amplification products (*i.e.*, "primer dimers") in later steps. Lack of 5'-phosphate termini also prevents ligation of multiple adapters to a protected fragment, insuring that a single adapter is ligated to each end of a protected fragment.

In order to catalyze ligation of two nucleic acid strands, most DNA ligases require the juxtaposition of a 5'-phosphate termini to a 3'-OH terminus. As a result of the lack of 5'-phosphate termini on the adapter oligonucleotides, when a duplex adapter oligonucleotide is joined to one end of a duplex protected fragment, ligation (*i.e.*, covalent joining by a ligase enzyme) will occur only between the strand comprising the 5'-phosphate terminated end of the protected fragment and the strand comprising the 3'-OH end of the adapter. However, the complementary strands, in which a 3'-OH end of the protected fragment is juxtaposed to a 5'-OH end of the adapter oligonucleotide, will not be ligated. Since this occurs at both ends of the protected fragment, the result of adapter ligation is a duplex, comprising a protected fragment flanked by adapters, with a nick near each 3'-end at the boundary between protected fragment sequences and adapter sequences.

The 3'-end of the "outer" end of the adapter(s) may be modified with blocking moieties, such as a 3' deoxynucleotide, 3' dideoxynucleotide, inverted nucleotide or hapten, to prevent wrong orientation ligation. However, as will be understood, even without such steps, some fraction of the products will be properly ligated, which is all that is necessary to achieve the final product. Without such modification, the efficiency may be affected which may reduce the utility in some applications, but it is not essential.

Covalent closure of the non-ligated adapter strand can be achieved by any method known in the art. In a preferred embodiment, the non-ligated adapter strand

is degraded by a nuclease, and resynthesized by a DNA polymerase. In one aspect, the nuclease is the T7 gene 6 exonuclease, and the polymerase is T4 DNA polymerase. Degradation and resynthesis can be sequential or simultaneous. In certain embodiments of the invention, nucleolytic and polymerization activity are 5 present in the same polypeptide. For example, the combined polymerase and 5'->3' exonuclease activities of a DNA polymerase (such as the Klenow fragment of *E. coli* DNA Polymerase I) are utilized to close the gap by "nick translation."

### **Amplification**

10 Duplex polynucleotides comprising a protected fragment flanked by adapter oligonucleotide duplexes are subjected to amplification.

### **Formation of tandem arrays of protected fragments**

15 Duplex polynucleotides comprising a protected fragment flanked by adapter oligonucleotide duplexes, or optionally their amplification products, are ligated into tandem arrays (concatemers) and the nucleotide sequence of the concatemer is determined, thereby providing the nucleotide sequences of the genetic alterations in the sample polynucleotides.

20 In order to efficiently generate tandem arrays, an internal subfragment containing a portion of the adapter ligated to the protected fragment, or its amplification product, and bearing a palindromic single-stranded overhang ("sticky end") may be generated by providing a restriction site within the adapter sequence and digesting the adapted products with a restriction enzyme prior to ligating the array. To prevent religation of the terminal ("outer") subfragments, a capture moiety 25 such as biotin may be provided on the terminal fragment by inclusion in the adapter and/or primer oligonucleotides, and removing them by capture with a recognition moiety such as streptavidin prior to ligation.

30 As will be understood, such ligation will generate arrays of highly random size, many of which will circularize, and thus be prevented from ligating into a vector. This effect will reduce the frequency of clones with inserts of appropriate size for sequencing. This may be controlled by including in the adapter population a

second set of adapters with a restriction site that creates the same sticky end as the bulk of the adapters, but which is cleaved exclusively by a second enzyme, by virtue of having difference adjacent nucleotides. One such example is BamH1 and BglI, both of which leave a 5'-GATC overhang, but recognize that sequence in the context 5 of differing adjacent nucleotides.

After cleaving the adapted products with BamH1, capturing the terminal portions, and ligating the internal fragments, the product can then be cleaved with BglI to generate linear arrays of more defined size. The average size of the arrays, or 10 number of inserts, may be adjusted by varying the ratio of BamH1-site containing adapters or primers to BglI-site containing adapters or primers.

An alternative means to generate defined arrays may be provided by including a small amount of a "chain-terminating" adapter oligonucleotide in the ligation reaction. Such molecules comprise double-stranded oligonucleotides bearing 15 at one end, a sequence complementary to the sticky ends of the adapted protected fragments, and on the other end, a single-stranded non-palindromic sequence that lacks the ability to form a hybrid with itself, either by Watson-Crick or other base-pairing that would otherwise lead to ligation of a dimer. This is most easily achieved by utilizing a non-palindromic sequence that cannot pair with itself. Inclusion of this chain terminator in the ligation reaction forces the formation of linear products 20 ending with the chain terminator adapters. A vector to which adapters complementary to the chain terminator sticky ends have been previously ligated ("decorated vector"), may then be added to the reaction to accept the arrays without allowing the closure of the vector or arrays into circular products.

Further, it will be recognized that the protected fragments or their amplified 25 products will contain mismatches, that when introduced into host cells, will give rise to mismatch correction mechanisms, including the post DNA replication repair system. This may be avoided by methylation of the ligation product with a DNA methylase corresponding to the host cell. Alternatively, one strand of the vector containing the ligated array may be removed by treatment with an exonuclease prior 30 to transformation. A suitable exonuclease is T7 gene 6 exonuclease. A single nick in the ligated product is required to prevent undesirable degradation of the vector. This

may be conveniently achieved by placing a non-ligating adapter on one end of the vector or the insert array, or by omitting the 5' terminal phosphate on the adapter. Alternatively, a vector may be used that contains a M13 or f1 origin of replication and cleaving it in one strand of the origin with gplI protein prior to treatment with T7 5 gene 6 exonuclease.

Finally, certain sequences may be subject to host restriction-modification systems. This may be overcome by appropriate host selection or pretreatment with a modification methylase.

10 **Analysis of sequence of tandem arrays**

DNA from colonies containing tandem arrays is prepared and sequenced using methods well understood in the art.

In a preferred embodiment, all of the reactions following purification of the protected fragments are performed sequentially in the same reaction vessel, by 15 sequential addition of enzymes (and buffers) required for the subsequent step(s).

The following examples are intended to illustrate, but not limit the invention.

**EXAMPLES**

20 **Example 1. Detection of mutations in cloned mutS genes.**

Five circular plasmid clones of *E. coli* mutS containing one mutation each in a 2kb region of the mutS coding sequence (each approximately 6kb total including a vector sequence) were linearized with EcoRI. Then 0.1 pmol (0.42 $\mu$ g) aliquots of each were combined in 25 $\mu$ l of 2.5M guanidine isothiocyanate (GTC) and 25 heteroduplexed by heating to 95°C for 2 min. and incubating at 37°C for 1h. The GTC was then removed by chromatography on a Sephadex G-25 spin column (AP Biotech).

Two  $\mu$ g of the purified heteroduplex was mixed with 10 pmol of *E. coli* mutS protein (AP Biotech) in 20 $\mu$ l 50mM Tris pH7.5, 8 mM MgCl<sub>2</sub>, and 0.5 mM DTT and 30 incubated for 1h. on ice. Then, 1 $\mu$ l 2 mg/ml DNase I was added and the sample incubated for 10 min. at 37°C. The digestion was terminated by addition of 23 $\mu$ l 50

mM Tris pH7.5, 10mM EDTA, and the mutS-protected complexes purified by chromatography on a Sephadryl S-200 spin column (AP Biotech). Twenty  $\mu$ l of the eluate was adjusted to 0.1M Tris pH7.5, 5 mM MgCl<sub>2</sub>, 7.5mM DTT, 100 $\mu$ M ATP and 1 mM each dATP, dGTP, dCTP, and TTP. The termini of the protected 5 fragments were then polished by the addition of 10 U T4 DNA polymerase and incubation for 10 min. at 37°C. The products were then adapted by addition of 35 pmol each forward and reverse double-stranded adapters (annealed sequences 1 and 2 mixed with annealed sequences 3 and 4, each separately pretreated with calf intestinal phosphatase (CIP)) with 2 units T4 DNA ligase (Stratagene) and incubation 10 for 1h. at 16°C. The unligated strands of the adapters (corresponding to sequences 2 and 4) in the ligation products (10 $\mu$ l) were then removed and the 3' ends of the protected fragments simultaneously extended by addition of 50 U T7 gene 6 exonuclease (AP Biotech) and incubation for 30 min. at room temperature. To 15 reduce background arising from residual adapter ligation, the final products were digested with 10 units Dra I in the same buffer for 1h. at 37°C.

The final products were amplified by denaturing for 1 min. at 95°C, followed by 29 cycles of PCR, with denaturation at 95°C for 10 seconds, primer annealing at 62°C for 20s, and primer extension at 72°C for 10s, followed by a final extension for 5 min. at 72°C using 1 $\mu$ M oligonucleotides 1 and 3 in a 50 $\mu$ l reaction containing 2.5 20 U Taq DNA polymerase. The products were digested with 80 U Hind III (New England Biolabs) for 3h. at 37°C. The released biotinylated adapter termini were removed after adding NaCl to 0.7 M by addition 40 $\mu$ l prewashed streptavidin agarose (Gibco-BRL) and incubation for 30 min. at room temperature. The mixture was extracted with phenol/chloroform and desalting by chromatography on a G-50 spin 25 column. A portion (7  $\mu$ l) of the eluate was then ligated with 100 ng Hind III-digested and phosphatase-treated pUC19 with 2 U T4 DNA ligase overnight at 14°C in a final volume of 10 $\mu$ l. The products were used to transform competent cells which were plated on LB/ampicillin agar. Analysis indicated 50% of the clones contained 30 inserts. All of the insert-containing clones contained single protected fragments.

Among 64 clones, 3 separate isolates were obtained for each of three of the mutation sites. The sequence changes corresponding to the mutations were properly

identified in each by sequence variation between the isolates. Nine of the clones contained inserts corresponding to the fourth site, and five corresponded to the fifth site. Seven other clones had inserts to a site (and indicated a mutation by sequence variation) which was not contained in the sequenced region containing the known mutations. Four clones contained sequences within a 40 nucleotide region in mutS but which did not identify any nucleotide changes by sequence variation. Nine of the clones corresponded to vector sequence also without changes from the known sequence. Seventeen clones either contained very small (<10 bp) inserts or inserts which matched neither mutS nor vector sequence; several of these clones had homology to prokaryotic sequences suggesting contaminating *E. coli* genomic DNA in the plasmid preparations.

**Example 2. Detection of single-nucleotide polymorphisms in a 4kb PCR product from APC using a solid-phase mutS protection reaction.**

Genomic DNA samples from four normal individuals were combined in equal mass ratio. A portion of exon 15 of the adenomatous polyposis coli (APC) gene was amplified from 400ng of the pooled genomic template DNA using 1 $\mu$ M oligonucleotide sequences 5 and 6 as primers and Pfu Turbo polymerase (Stratagene) in the supplied buffer supplemented with 5% glycerol. The reaction was heated to 95°C for one minute, followed by 29 cycles of 95°C denaturation for 1 min., 62°C primer annealing for 1 min. and extension at 72° for 6 min., followed by a final extension at 72°C for 10 minutes, generating a product ~4kb in length.

A reference DNA product was prepared in parallel except using 15 ng of a cloned PCR product from one of the four individuals as template and biotinylated primers corresponding to oligonucleotides 5 and 6.

Biotinylated reference DNA (0.5 $\mu$ g) and patient DNA (0.5 $\mu$ g) were heteroduplexed in 150 $\mu$ l 50% formamide, 2.78X SSPE and heated to 95°C for 5 minutes followed by incubation at 37°C for 1h. The buffer was exchanged and the sample concentrated by diafiltration on a Centricon 100 followed by a single wash with 2ml 10mM Tris, 1mM EDTA pH7.5 (TE). The product was diluted to 200 $\mu$ l with 1X B&W (1M NaCl, 10mM Tris, 1mM EDTA, 0.1% Tween 20), and 100 $\mu$ g

Dynal M-280 Streptavidin (Dynal Corp., prewashed with 1X B&W) was added and the hybrids captured by end-over-end mixing for 30 min. at room temperature. The particles were collected on a magnet, washed once with 300 $\mu$ l mutS binding buffer (50mM Tris pH7.5, 8 mM MgCl<sub>2</sub>, and 0.5mM DTT), and resuspended in 20 $\mu$ l of the same buffer. MutS (10 pmol) was added and the suspension incubated on ice without mixing for 1 hr. DNase I (2 $\mu$ g in 1 $\mu$ l) was then added and the suspension incubated for 10 minutes at 37°C. The digest was terminated by addition of 23 $\mu$ l 50mM Tris pH7.5, 10mM EDTA. The particles were removed by applying a magnetic field, and the supernatant was chromatographed over a Sephacryl S-200 MicroSpin column (AP Biotech).

A portion of the S-200 eluate (20 $\mu$ l) was processed as described above (example 1), and the Hind-III digested inserts ligated into restricted and phosphatased pUC19. The cloned products were sequenced. Thirty one of the 45 inserts sequenced corresponded to a single (A/G) polymorphism at position 5034 of the gene (codon 1678). One insert corresponded to the (G/A) polymorphism at position 4479 (codon 1493) and two inserts corresponded to the (A/G) polymorphism at position 5880 (codon 1960). Eleven inserts contained APC sequence but did not show any changes from the expected sequence.

**20 Example 3. Generation of cloned tandem protected-fragment insert arrays.**

A pUC vector decorated with nonpalindromic adapter ends is generated. One hundred micrograms of pUC19 is digested with BamHI and EcoRI, and the large fragment separated on a 1% agarose gel and collected by electroelution. The vector 25 is ligated with two oligonucleotide adapters (Q and R) created by mixing and annealing sequence 7 with sequence 8 (Q) and sequence 9 with sequence 10 (R).

After removal of the adapter termini from the amplified mutS-protected fragments by HindIII digestion and capture on streptavidin agarose as described in example 1, the products are quantitated either by absorbance or fluorescence using 30 ethidium bromide or Hoescht 33258. The mixture is supplemented with 5% (mol/mol, assuming an average insert + adapter remnant size of 33bp) of adapter "S"

created by annealing oligonucleotide sequences 11 and 12. This adapter provides a HindIII-compatible single-stranded terminus and a sequence complementary to the adapter-modified ends of the pUC19 vector above. Alternatively, an adapter "S2" is created by annealing oligonucleotide sequences 11 and 17, and may be used to limit  
5 the formation of S2 dimers during the subsequent ligation reaction. The mixture is ligated overnight at 16°C with 2 U T4 DNA ligase (Stratagene) in 50mM Tris pH7.5, 5mM MgCl<sub>2</sub>, 1mM DTT and 15% polyethylene glycol (Sigma) to create a tandem array terminated in non-self-ligatable 5'-GCTA-3' single-stranded overhangs. The arrays are then ligated into the decorated vector by addition of 1µg of the pUC19  
10 prepared as described above and incubation for 4h. at 16°C. The ligase is then inactivated by heating the mixture to 65°C for 15 min. To minimize the possibility of intracellular mismatch repair on cross-hybridized, mismatched PCR products in the final ligation product, one strand of the vector-ligated product may be removed by addition of 100 units T7 gene 6 exonuclease (AP Biotech) and incubation at 30°C  
15 for 20 min. A portion (1µl) of the final reaction product diluted 1:10 in 10mM Tris, 1 mM EDTA is used to transform 25µl of MAX efficiency DH5α *E. coli* (Gibco-BRL). Colonies are screened for a preferred insert size of 300-800 bp by minipreparation or colony PCR. Background arising from occasional ligation of the "S" adapters may be eliminated by digestion of the final (double-stranded) ligation  
20 product with Sac I prior to transformation (or T6 gene 6 exonuclease treatment, if used).

**Example 4. Removal of inserts arising from polymorphisms in APC.**

A set of amplified, protected fragments from sites of single-nucleotide  
25 polymorphisms (SNP amplimers) in a gene of interest is generated as in example 2 from a pool of normal DNA, except using different adapters (and corresponding primers; adapter X = oligonucleotide sequences 13 and 14, sequence 13 as primer; adapter Y = oligonucleotide sequences 15 and 16, sequence 15 as primer). Any such sequences may be used as long as they function as adapter/primers in the method  
30 presented in examples 1 and 2 and will not anneal to the adapter sequences used for processing patient DNA. The primers for generating SNP amplimers have

biotinylated termini allowing their removal by capture on a streptavidin solid support. A set of amplimer products generated from patient DNA using the adapters and primers as in example 2, except lacking biotin modifications is heated in the presence of a 10-fold molar excess of biotinylated SNP amplimers. Two  $\mu$ l of the

5 PCR reaction from patient DNA is combined with 20 $\mu$ l of a PCR reaction containing the SNP amplimers in addition to 1.5M sodium thiocyanate, 120mM disodium phosphate, 10mM EDTA in a final volume of 100 $\mu$ l. To minimize cross-hybridization of the amplimers due to the adapter sequences used for the SNP and patient DNAs, the annealing is supplemented with 50 pmol each oligonucleotide

10 sequences 2, 4, 14 and 16. To this mixture, 8 $\mu$ l redistilled phenol is added and the mixture heated to 100°C for 10 min., and then chilled on ice. The reactions are then placed in a thermocycler and cycled for 2 min. at 65°C and 15 min. at 25°C for a total of 10 cycles (~2.5h) (Miller, R, and Riblet, R., *Nucl. Acids Res.* 23:2339-2340 (1995)). The mixture is extracted once with chloroform and 50 $\mu$ l is desalted over a

15 G50 spin column (AP Biotech). The eluate is then mixed with 50 $\mu$ l 2M NaCl, and incubated with 50 $\mu$ l streptavidin agarose (Gibco-BRL) equilibrated with 1M NaCl, 10mM Tris pH7.5, 1mM EDTA with mixing for 30 min. The final product is then desalted by chromatography on a G-50 spin column. Two  $\mu$ l of the G-50 eluate is reamplified with Taq DNA polymerase using biotinylated primers (sequences 1, 4)

20 by heating to 95°C for 1 min., and then cycling at 95°C for 10 sec., annealing at 62°C for 20 sec. and extending at 72°C for 10 sec. for a total of 14 cycles. The product(s) are digested with Hind III and processed as described above (examples 2 and 3) to generate tandem insert array clones.

It is to be understood that while the invention has been described in conjunction with the above embodiments, that the foregoing description and examples are intended to illustrate and not limit the scope of the invention. Other aspects, advantages and modifications within the scope of the invention will be apparent to those skilled in the art to which the invention pertains.

**Sequences (Referenced above)****Oligonucleotide #**

1. 5'-BIO-TsGsCs-TsAC-CAG-TGC-CAG-CCA-AGC-TTT-T-3'  
5 where s = phosphorothioate linkage
2. 5'-AAA-AGC-TTG-GCT-GGC-ACT-GGT-AGC-AZ-3'  
where Z = 3' inverted A (3'-3' linkage)
3. 5'-BIO-CsCsT-sCsAA-GGA-TGG-CTC-CGA-AGC-TTT-T-3'  
where s = phosphorothioate linkage
4. 5'-AAA-AGC-TTC-GGA-GCC-ATC-CTT-GAG-GZ-3'  
10 where Z = 3' inverted A (3'-3' linkage)
5. 5'-GTT-GAA-CTC-TGG-AAG-GCA-AAG-TCC-T-3'
6. 5'-TTT-CTA-CCA-GGG-GAA-ATT-GAG-TTT-3'
7. 5'-pTAG-CCG-AGG-GC-3'  
15 where p = 5' phosphate
8. 5'-pGAT-CGC-CCT-CG-3'
9. 5'-pAAT-TCC-GCC-TG-3'
10. 5'-TAG-CCA-GGC-GG-3' (5'-OH)
11. 5'-pGC-TAG-GGC-CG-3'
12. 5'-pAG-CTC-GGC-CC-3'  
20
13. 5'-BIO-TsGsA-sCsGT-GCA-CTC-GGG-CGG-GAT-CCT-T-3'  
where s = phosphorothioate linkage
14. 5'-AAG-GAT-CCC-GCC-CGA-GTG-CAC-GTC-AZ-3'  
where Z = 3' inverted A (3'-3' linkage)
- 25 15. 5'-BIO-TsGsA-sCsGC-GCT-GCC-ATG-CCG-GAT-CCT-T-3'  
where s = phosphorothioate linkage
16. 5'-AAG-GAT-CCG-GCA-TGG-CAG-CGC-GTC-AZ-3'  
where Z = 3' inverted A (3'-3' linkage)
17. 5'-pAG-TTC-GGC-CC-3'

CLAIMS

What is claimed is:

1. A method for determining the sequence of a sample polynucleotide containing one or more genetic alterations, wherein the method comprises:
  - 5 (a) contacting a heteroduplex containing a sample polynucleotide and a reference polynucleotide with a reagent which recognizes a non-duplex polynucleotide structure, to form a reagent-heteroduplex complex, wherein the duplex contains at least one base-pair mismatch;
  - (b) contacting the reagent-heteroduplex complex with a nucleolytic agent
  - 10 to produce a protected fragment;
  - (c) joining a double-stranded adapter oligonucleotide to the protected fragment;
  - (d) amplifying the products of step (c) using primers complementary to a portion of the adapter oligonucleotide, to generate amplification products;
  - 15 (e) joining the amplification products to one another to form a concatemer, wherein each monomeric unit of the concatemer comprises a region of the sample polynucleotide containing the genetic alteration, and the monomeric units are separated by regions of sequence corresponding to a portion of the adapter oligonucleotide sequence; and
  - 20 (f) determining the nucleotide sequence of the concatemer.
2. The method according to claim 1, wherein a plurality of reference polynucleotides are used.
- 25 3. The method according to claim 1, further comprising step (b)(i), wherein the ends of the protected fragments are repaired.
4. The method according to claim 1, wherein the adapter oligonucleotide comprises one or more restriction enzyme recognition sites.

30

5. The method according to claim 4, further comprising step (d)(i), wherein the amplification products are digested with a restriction enzyme which cleaves at the recognition site.

6. The method according to claim 1, wherein the adapter  
5 oligonucleotide:

- (a) comprises an inner end and an outer end,
- (b) is non-phosphorylated at the 5' terminus of the inner end,
- (c) comprises a capture moiety at the outer end, and
- (d) comprises one or more blocking linkages adjacent to the capture

10 moiety.

7. The method according to claim 1 wherein, in step (c), two adapter oligonucleotides, having different sequences, are used.

8. The method according to claim 1, wherein, in step (c), joining the adapter comprises the following steps:

15 (c)(i) joining the first strand of the double-stranded adapter oligonucleotide to the products of step (b), and  
(c)(ii) joining the second strand of the double-stranded adapter oligonucleotide to the products of step (c)(i).

9. The method according to claim 8, wherein  
20 (a) the first strand is covalently joined to the protected fragment by ligation, and  
(b) the second strand is covalently joined to the protected fragment by nick-translation.

10. A method for determining the sequences of one or more genetic  
25 alterations in a plurality of sample polynucleotides, wherein the method comprises:  
(a) contacting a plurality of duplexes with a reagent which recognizes a non-duplex polynucleotide structure, wherein each duplex comprises a sample polynucleotide strand and a reference polynucleotide strand and at least one of the duplexes contains at least one base-pair mismatch, to form at least one reagent-  
30 heteroduplex complex;

- (b) contacting the reagent-heteroduplex complexes with a nucleolytic agent to produce a plurality of protected fragments;
- (c) joining a double-stranded adapter oligonucleotide to the protected fragments;

5 (d) amplifying the products of step (c) using primers complementary to a portion of the adapter oligonucleotide, to generate a plurality of amplification products;

- (e) joining the amplification products to one another to form a concatemer, wherein each monomeric unit of the concatemer comprises a region of a sample polynucleotide containing a genetic alteration, and the monomeric units are separated by regions of sequence corresponding to a portion of the adapter oligonucleotide sequence; and
- (f) determining the nucleotide sequence of the concatemer.

10 11. The method according to claim 10, wherein a plurality of reference polynucleotides are used.

12. The method according to claim 10, further comprising step (b)(i), wherein the ends of the protected fragments are repaired.

13. The method according to claim 10, wherein the adapter oligonucleotide comprises one or more restriction enzyme recognition sites.

14. The method according to claim 13 further comprising step (d)(i), wherein the amplification products are digested with a restriction enzyme which cleaves at the recognition site.

15. The method according to claim 10, wherein the adapter oligonucleotide: (a) comprises an inner end and an outer end, (b) is non-phosphorylated at the 5' terminus of the inner end, (c) comprises a capture moiety at the outer end, and (d) comprises one or more blocking linkages adjacent to the capture moiety.

16. The method according to claim 10 wherein, in step (c), two adapter oligonucleotides, having different sequences, are used.

30 17. The method according to claim 10, wherein, in step (c), joining the adapter comprises the following steps:

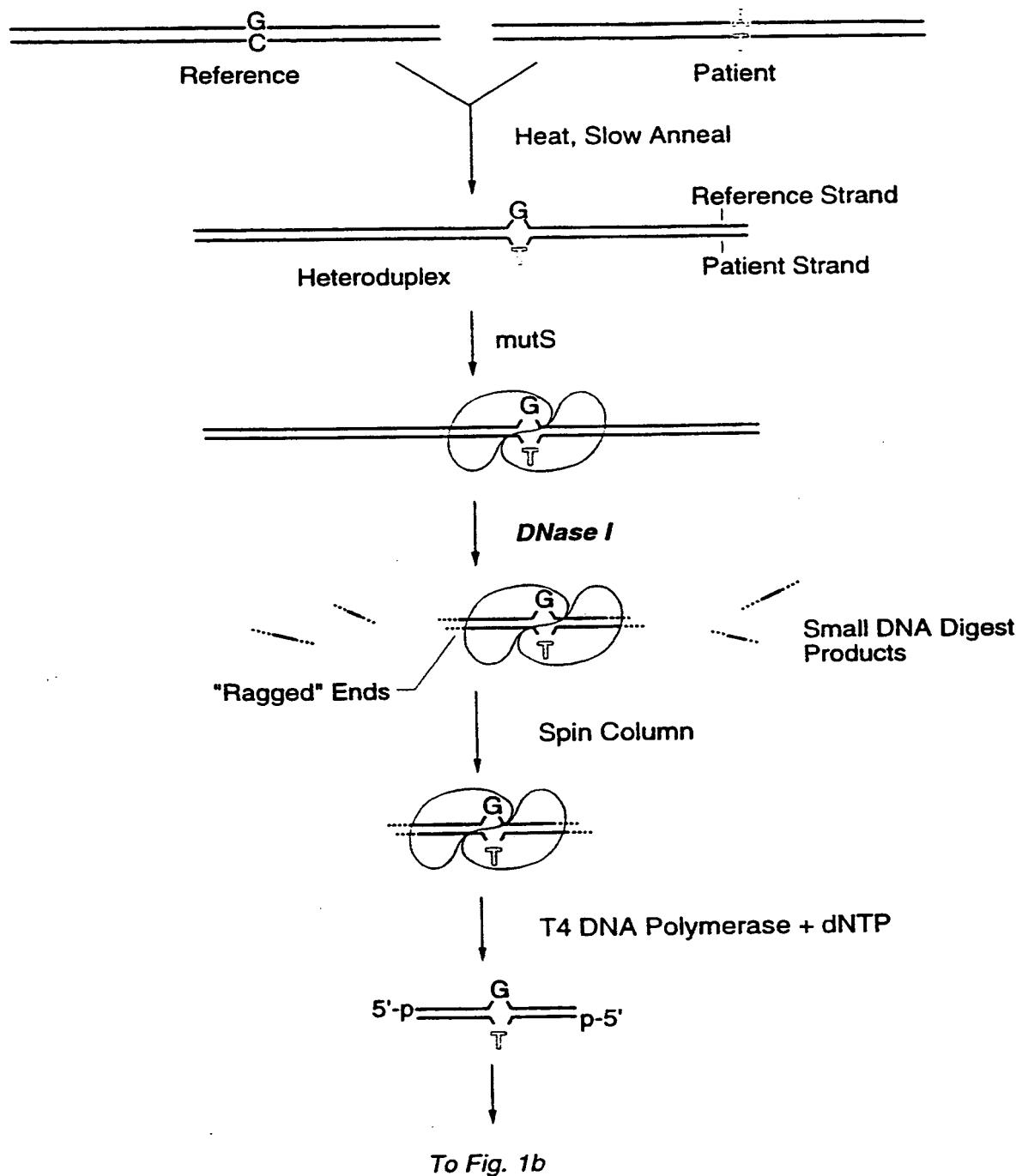
(c)(i) joining the first strand of the double-stranded adapter oligonucleotide to the products of step (b), and

(c)(ii) joining the second strand of the double-stranded adapter oligonucleotide to the products of step (c)(i).

5 18. The method of claims 1 or 10, wherein the joining of steps (c) and (e) comprises the formation of a covalent bond.

19. The method of claim 8 or 17, wherein the joining comprises the formation of a covalent bond.

FIG. 1A



From Fig. 1A

FIG. 1B

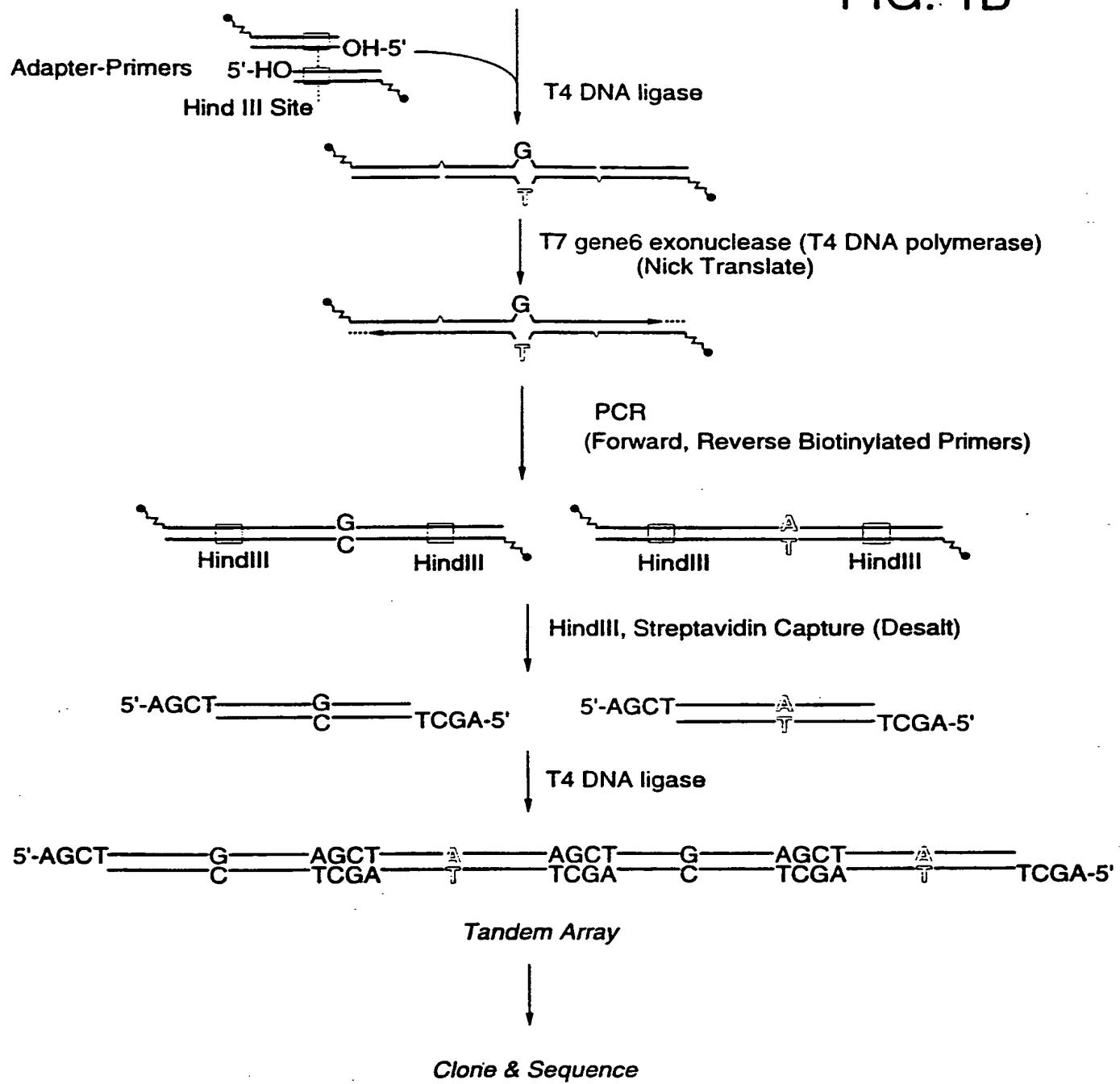
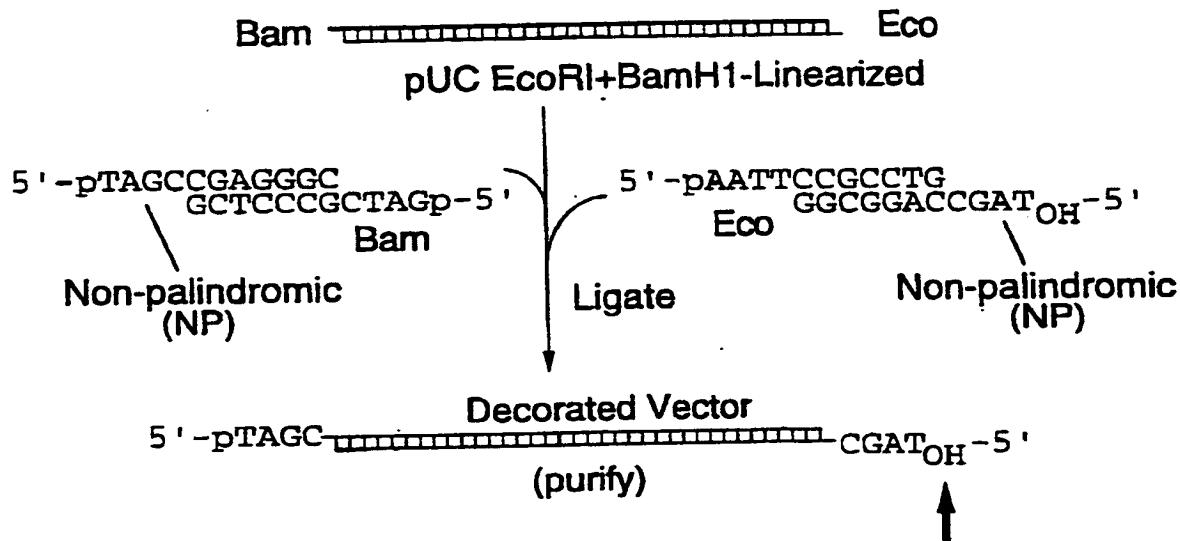
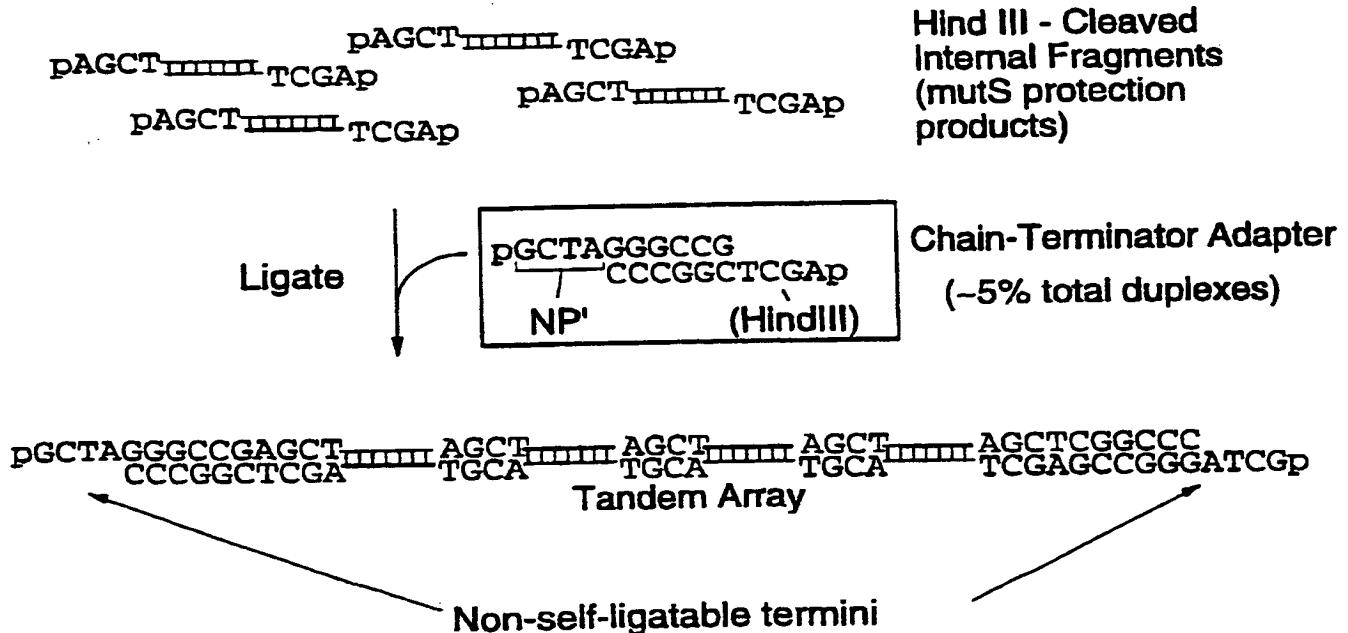
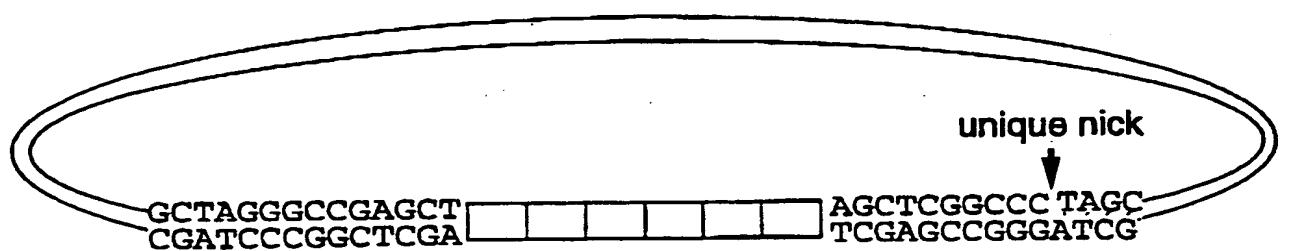
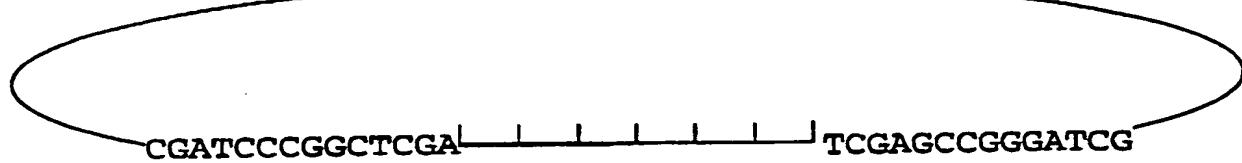


FIG. 2AB

**A. Decorated Vector Construction****B. Tandem Insert Generation**

**FIG. 2C****C. Clone Construction****Decorated Vector (above) + Tandem Array****Ligate (exhaustive)****T7 gene 6 Exonuclease****Single-stranded pseudo-phagemid**

THIS PAGE BLANK (USPTO)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
8 February 2001 (08.02.2001)

PCT

(10) International Publication Number  
WO 01/09384 A3

(51) International Patent Classification<sup>7</sup>: C12Q 1/68

(21) International Application Number: PCT/US00/20557

(22) International Filing Date: 28 July 2000 (28.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/146,437 29 July 1999 (29.07.1999) US

(71) Applicant (for all designated States except US): GENZYME CORPORATION [US/US]; Legal Department, Metrowest Place, 4th floor, 15 Pleasant Street Connector, Framingham, MA 01701-9322 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): STEFANO, James, E. [US/US]; 199 Saddle Hill Road, Hopkinton, MA 01748 (US).

(74) Agents: KONSKI, Antoinette, F. et al.; Baker & McKenzie, 660 Hansen Way, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— With international search report.

(88) Date of publication of the international search report:  
7 June 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/09384 A3

(54) Title: SERIAL ANALYSIS OF GENETIC ALTERATIONS

(57) Abstract: Methods and compositions for the rapid detection and identification of multiple genetic alterations are provided.

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/20557

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, PAJ, BIOSIS, MEDLINE, EMBASE, CHEM ABS Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>PARSONS B L ET AL: "EVALUATION OF MUTS AS A TOOL FOR DIRECT MEASUREMENT OF POINT MUTATIONS IN GENOMIC DNA" MUTATION RESEARCH, NL, AMSTERDAM, vol. 374, no. 2, 21 March 1997 (1997-03-21), pages 277-285, XP002914507 ISSN: 0027-5107 the whole document</p> <p>---</p> <p style="text-align: center;">-/-</p>	1-19



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*Y\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*&\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*&\* document member of the same patent family

Date of the actual completion of the international search

28 February 2001

Date of mailing of the international search report

06/03/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Hagenmaier, S

# INTERNATIONAL SEARCH REPORT

Int. Application No

PCT/US 00/20557

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ELLIS L A ET AL: "MUTS BINDING PROTECTS HETERODUPLEX DNA FROM EXONUCLEASE DIGESTION IN VITRO: A SIMPLE METHOD FOR DETECTING MUTATIONS" NUCLEIC ACIDS RESEARCH, GB, OXFORD UNIVERSITY PRESS, SURREY, vol. 22, no. 13, 11 July 1994 (1994-07-11), pages 2710-2711, XP000606262 ISSN: 0305-1048 the whole document ---	1-19
Y	VELCULESCU V E ET AL: "SERIAL ANALYSIS OF GENE EXPRESSION" SCIENCE, US, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, vol. 270, 20 October 1995 (1995-10-20), pages 484-487, XP002053721 ISSN: 0036-8075 the whole document ---	1-19
Y	EP 0 761 822 A (UNIV JOHNS HOPKINS MED) 12 March 1997 (1997-03-12) the whole document ---	1-19
A	WO 95 29258 A (ST JAMES AND SEACROFT UNIVERSI ;TAYLOR GRAHAM ROY (GB)) 2 November 1995 (1995-11-02) the whole document ---	
A	WO 97 29211 A (US HEALTH ;WEINSTEIN JOHN N (US); BOULAMWINI JOHN (US)) 14 August 1997 (1997-08-14) the whole document ---	
A	WHITE M J ET AL: "CONCATEMER CHAIN REACTION: A TAA DNA POLYMERASE-MEDIATED MECHANISM FOR GENERATING LONG TANDEMLY REPETITIVE DNA SEQUENCES" ANALYTICAL BIOCHEMISTRY, US, ACADEMIC PRESS, SAN DIEGO, CA, vol. 199, no. 2, 1 December 1991 (1991-12-01), pages 184-190, XP000236491 ISSN: 0003-2697 the whole document ---	
A	WO 99 23256 A (LUCITO ROBERT ;WIGLER MICHAEL (US); COLD SPRING HARBOR LAB (US)) 14 May 1999 (1999-05-14) the whole document ---	
A	US 5 750 335 A (GIFFORD DAVID K) 12 May 1998 (1998-05-12) the whole document ---	

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/20557

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 41002 A (GENZYME CORP) 19 December 1996 (1996-12-19) page 5 -page 6, line 14 ---	
A	COTTIN R: "Current methods of mutation research" MUTATION RESEARCH, vol. 285, 1993, pages 125-144, XP000443992 page 132, paragraph 4 -page 133 ---	
P, Y	WO 99 39003 A (GENZYME CORP) 5 August 1999 (1999-08-05) the whole document -----	1-19

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte. lational Application No

PCT/US 00/20557

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
EP 0761822	A 12-03-1997	US 5695937 A			09-12-1997
		US 5866330 A			02-02-1999
		AU 707846 B			22-07-1999
		AU 6561496 A			20-03-1997
		AU 7018896 A			01-04-1997
		CA 2185379 A			13-03-1997
		DE 761822 T			11-01-2001
		GB 2305241 A, B			02-04-1997
		IE 80465 B			12-08-1998
		JP 10511002 T			27-10-1998
		WO 9710363 A			20-03-1997
WO 9529258	A 02-11-1995	AU 2315195 A			16-11-1995
		EP 0758402 A			19-02-1997
		US 5919623 A			06-07-1999
WO 9729211	A 14-08-1997	AU 2264197 A			28-08-1997
WO 9923256	A 14-05-1999	AU 1449799 A			24-05-1999
		EP 1032705 A			06-09-2000
US 5750335	A 12-05-1998	WO 9322457 A			11-11-1993
WO 9641002	A 19-12-1996	US 5571676 A			05-11-1996
		US 5707806 A			13-01-1998
		AU 705257 B			20-05-1999
		AU 5981396 A			30-12-1996
		CA 2219933 A			19-12-1996
		EP 0832283 A			01-04-1998
		JP 11506937 T			22-06-1999
WO 9939003	A 05-08-1999	AU 2005199 A			16-08-1999
		EP 1051512 A			15-11-2000

**THIS PAGE BLANK (USPTO)**